

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
16 January 2003 (16.01.2003)

PCT

(10) International Publication Number
WO 03/005232 A2

(51) International Patent Classification⁷: **G06F 17/30**

(21) International Application Number: PCT/CA02/01044

(22) International Filing Date: 8 July 2002 (08.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/303,036 6 July 2001 (06.07.2001) US

(71) Applicant (for all designated States except US): **ANGOSS SOFTWARE CORPORATION** [CA/CA]; 34 St. Patrick Street, Suite 200, Toronto, Ontario M5T 1V1 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **APPS, Eric** [CA/CA]; 34 St. Patrick Street, Suite 200, Toronto, Ontario M5T 1V1 (CA). **ONO, Ken** [CA/CA]; 34 St. Patrick Street, Suite 200, Toronto, Ontario M5T 1V1 (CA).

(74) Agent: **FASKEN MARTINEAU DUMOULIN LLP**; Toronto Dominion Bank Tower, Suite 4200, P.O. Box 20, Toronto-Dominion Centre, Toronto, Ontario M5K 1N6 (CA).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

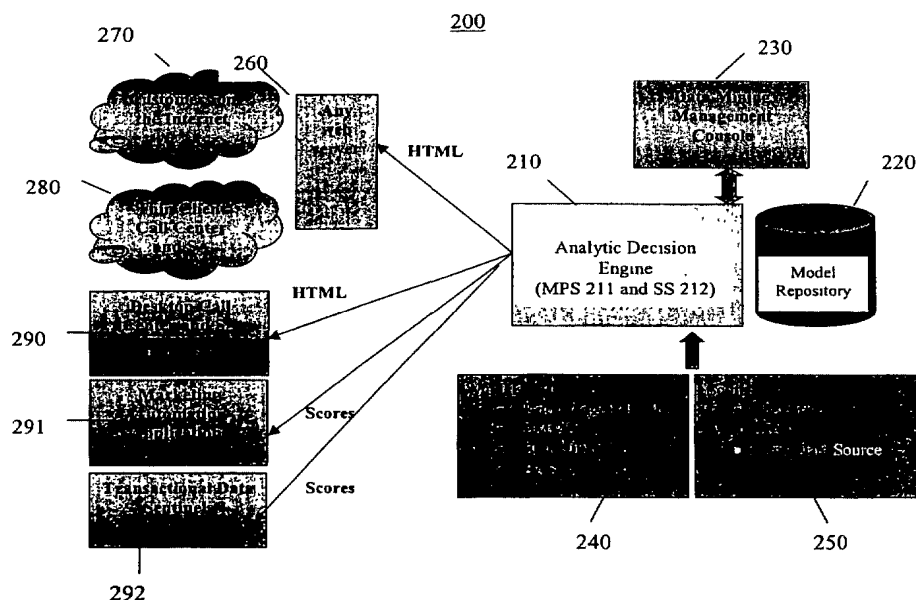
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: A METHOD AND SYSTEM FOR THE VISUAL PRESENTATION OF DATA MINING MODELS



(57) Abstract: A method for delivering presentations associated with data mining models. The method includes the steps of: a) storing the data mining models and the presentations in a repository in a data mining system; b) associating ones of the data mining models with ones of the presentations through references within ones of the presentations; c) receiving a request from a user for a presentation; d) processing the request to determine an appropriate presentation corresponding to the request; the request including a customer identification; the request including an event identification; and, e) delivering the appropriate presentation to the user.



WO 03/005232 A2



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

A METHOD AND SYSTEM FOR THE VISUAL PRESENTATION OF DATA MINING MODELS

The invention relates to the field of data mining systems, and more specifically to a system for the visual presentation of outcomes of data mining models.

5 BACKGROUND OF THE INVENTION

Digital Ecosystems and Data Creation. A growing number of interactions between individuals and organizations and among organizations (e.g. transactions, exchanges, communications, business processes, and other actions) are conducted through computing systems and across computing networks. The environment within which
10 each organization conducts these interactions with others (e.g. prospects, customers, intermediate product and service suppliers and other intermediaries across the value chain between the ultimate producer of a good or service and its ultimate consumer) can be thought of as its “digital ecosystem”.

These interactions are person-to-person (i.e. human aided), person-to-system (i.e.
15 partially automated), and system-to-system (i.e. fully automated). They generate resultant data. As digital ecosystems become more pervasive and interconnected, as the number of intermediaries connected to them grow, and as the numbers and types of interactions being conducted across them expand, many organizations have access to growing volumes of data derived from a variety of different sources.

20 *Need for Improved Analysis and Decisioning Systems.* This data is potentially useful to organizations that wish to better manage their business operations. Much of this data is largely unused while being stored in systems that are costly to maintain.

Prior Art Data Mining Systems. In this context, the concept of “data mining” has emerged commercially over the past decade as one means for organizations to analyze
25 data in order to extract useful knowledge therefrom. In general, data mining or knowledge discoveries involves the analysis of data structures for novel or unknown relationships that provide information that can be acted upon.

Current data mining systems are based on the use of knowledge discovery algorithms to automatically analyze data to discover relationships, patterns, knowledge or other

useful information. These data mining systems ("DM Systems") currently consist of data mining software tools incorporating data mining algorithms ("DM Tools"), and, more recently, relational database management systems ("RDMS") incorporating data mining algorithms, or enterprise software applications ("DM Applications")
5 containing embedded data mining algorithms.

Referring to FIG. 1, there is shown a block diagram illustrating a typical data mining system **100** in accordance with the prior art. Data created by person-to-person, person-to-system, and system to system interactions is created, exchanged and captured through the Internet, by means of autonomous and interconnected wired and wireless
10 networks, which support a wide range of appliances, applications, and interfaces (e.g. browsers, PCs, PDAs, telephones, etc.) which connect individuals and organizations. This data, which is typically stored in a relational database system ("RDBMS System") **110** or in a data mining application ("DM Application") **120**, is extracted by a proprietary data mining system **130** operated by a skilled analyst. The output of the
15 data mining process engaged in by such analyst, in the form of a predictive model or set of scores relevant to the business objective and data structure being analyzed **140** may be returned to the RDBMS System **110** or DM Application **120** through a batch process as static scores. The RDBMS System **110** typically includes proprietary data mining algorithms and supports the application of these algorithms to data stored in
20 the RDBMS System. Data mining models created by the RDBMS System are stored in the RDBMS as tables **112** and applications **111** can access these results if desired via a query or through an application programming interface ("API") with additional query formulation or coding. The DM Application **120** typically includes a data mining engine developed or licensed by the developer of the DM Application. Data
25 mining typically occurs on data stored in the application and resultant data mining models or scores are stored in the application for use in accordance with the internal, proprietary logic of the specific application (e.g. to support automated segmentation in a marketing application).

Current innovations in data mining systems have focused on the nature of data mining
30 algorithms, and the incorporation thereof into the DM System. The focus has been on the process of sourcing data for and creating data mining models, and the development of aids for the user to assess the results of or process of validating a data

mining model. For example, in United States Patent No. 6,112,194 (Bigus), discloses a feedback mechanism for monitoring performance of data mining tasks. The feedback mechanism consists of a user selected mining technique type for a data mining operation and an associated quality measure type. A quality indicator, calculated from the mining technique type and the quality measure type, is displayed to the user during the mining operation to help the user decide whether to stop the data mining operation and reconfigure the operation. Other patents related to data mining include United States Patent Nos. 6,226,648 (Appleman, et al.); 6,219,775 (Wade, et al.); 6,216,134 (Heckerman, et al.); 6,208,989 (Dockter, et al.); 6,205,472 (Gilmour); 6,192,356 (Eyles); 6,108,004 (Medl); 6,081,788 (Appleman, et al.); 6,055,510 (Henrick, et al.); 5,875,285 (Chang); and, 5,787,425 (Bigus). All of these patents describe various data mining techniques and algorithms and improvements thereto.

DM Tools (e.g. SAS Enterprise Miner™ and ANGOSS KnowledgeSTUDIO™) are software applications installed on a dedicated computer system for use by skilled analysts to perform data mining tasks (i.e. data exploration or knowledge discovery and predictive modeling). These DM tools enable users to import, analyze and model data sourced from a variety of files, database tables, or other data structures. RDBMS Systems (e.g. Microsoft SQL Server 2000 and Oracle 9i) incorporate data mining algorithms into the relational database environment, enabling the mining discovery of or relationships between data stored in tables contained in the relational database.

Thus a limitation of these systems is that rather than providing analyst interfaces, they typically provide an application programming interface that enables third parties to design and develop applications and interfaces integrated with these RDBMS Systems. DM Applications provide user interfaces and functionality enabling the application of data mining algorithms to data captured or accessible to the DM Application in a specific business domain. The data mining capabilities of DM Applications may be developed by the application provider (e.g. E.Piphany, Blue Martini, etc.), or through integration of the enterprise software application with a data mining engine incorporated into an RDBMS System, or using an integrated data mining engine that provides an application programming interface for this purpose (e.g. ANGOSS KnowledgeSERVER™ and KnowledgeSTUDIO™ SDK).

In summary, the focus of existing DM systems is to support the application of data mining algorithms to data structures to promote the development and validation of a data mining model. These DM systems are written in a proprietary format in a project centric fashion, for use within a proprietary application environment of the DM vendor, for deployment into the organization's production environment as a "score" through a batch process to a mainframe or RDBMS System.

As a result, the traditional approach to data mining has been project oriented and is designed for use by a relatively small group of skilled analysts. Their activities include defining a business objective; sourcing relevant data from different operational data sources for analysis; creating a data mining model based on this analysis; displaying the data mining model in some limited fashion in the form of a report (i.e. a visualization) or the creation of scores (i.e. a numeric representation) placed in a relational database or mainframe computer; and, measuring the validity and usefulness of the data mining model by reference to the defined business objective.

Prior art systems are deficient because of inefficiencies associated with the data mining process and the limited accessibility and usefulness of data mining models once produced. Data mining models are costly to create, requiring skilled personnel to operate and expensive systems to maintain, monitor and optimize; and, difficult to integrate with the systems and applications comprising the organization's digital ecosystem where access to such data mining models might be usefully deployed. Even when a data mining model (e.g. a customer profitability model, a transaction fraud model, or a next product suggestion model) has been created, it has proven to be extremely difficult to integrate the data mining model with other systems in the enterprise, where this knowledge might be more usefully applied, such as other applications, or applications operating in geographically remote locations or on different computing platforms.

Accordingly, it is difficult to create and move these data mining models (i.e. scores, rules, and rules systems) to heterogeneous operational environments or to a variety of different applications and business processes at the enterprise level. Similarly, even where a data mining score, rule, or rules system is moved to a production environment, it is difficult to fully implement data mining models which rely on a

variety of different data inputs from disparate data sources or to assess the performance of the data mining score, rule, or rules system with existing systems in the operational environment. Furthermore, it is difficult to integrate the results of a multitude of data mining models with other operational systems used to guide decision making. For example, it is difficult to integrate the results of a profitability model with credit limit approvals, new product offers or other similar business strategies and business processes relating to customers based on their profitability.

A need therefore exists for the effective deployment, management and optimization of a multitude of data mining models in a plurality of computing environments, and for the association of these data mining models with business strategies and business processes. Consequently, it is an object of the present invention to obviate or mitigate at least some of the above mentioned disadvantages.

SUMMARY OF THE INVENTION

The invention seeks to provide a solution to the problem of creation, storage and management of numerous data mining models, including the creation of templates for common data mining projects, interfaces for the building and deployment of data mining models, interfaces for managing a multitude of data mining models in heterogeneous computing environments, interfaces for the association of data mining models in both static and dynamic modes with presentations, and the continuous evaluation and optimization of these models and presentations in a plurality of computing environments.

In accordance with this invention there is provided, a method for delivering presentations associated with data mining models comprising the steps of: a) storing the data mining models and the presentations in a repository in a data mining system; b) associating ones of the data mining models with ones of the presentations through references within ones of the presentations; c) receiving a request from a user for a presentation; d) processing the request to determine an appropriate presentation corresponding to the request; and, e) delivering the appropriate presentation to the user.

In one instance the request includes a customer identification or an event identification. In another aspect of the method includes the steps of creating a database schema for the repository; creating data mining models and importing these into the data mining system; and, creating presentations and importing these into the
5 data mining system.

The method includes the additional steps of receiving inputs from users and from reporting systems included in the data mining system; updating the data mining models and presentations based on these inputs to produce updated data mining models and updated presentations; and, storing these updated data mining models, and
10 updated presentations in the repository. The method comprises the additional steps of defining rules under which the data mining models are associated with presentations; and, defining rules under which the presentations are selected and prepared for delivery.

According to another aspect of the invention, a method is provided wherein the data
15 mining models created can include simple and complex scores, static and dynamic scores, rules, rules sets, and rules systems, and composites and combinations of such scores, rules, rules sets, and rules systems. According to another aspect of the invention, a method is provided wherein the presentations include defined sets of content including data mining model references, scores, rules, attributes, activity
20 suggestions, to do lists, pop-up windows, HTML segments, XML scripts, and sets of computer instructions. According to another aspect of the invention, a method is provided wherein the presentations include simple presentations, range based presentations, and complex presentations.

According to another aspect of the invention, a method is provided wherein the method
25 steps are accomplished over networks including the Internet and networks based on Internet technologies. According to another aspect of the invention, a method is provided wherein users include data mining management consoles integrated with or connected over networks that facilitate all aspects of the data mining process from data access, through preparation and data mining model creation to model deployment
30 to disparate business user environments. According to another aspect of the invention, a method is provided wherein users include applications running on a computer, system, or other appliance integrated with or connected over networks. According to

another aspect of the invention, a method is provided wherein users include thin or thick clients in a client-server or browser-server environment integrated with or connected over networks. According to another aspect of the invention, a method is provided wherein the reporting systems within the data mining system include outcome monitoring systems and presentation usage monitoring systems.

According to another aspect of the invention, a method is provided wherein second inputs include updates, edits, modifications, deletions, replacements, suspensions, changes, improvements, and optimizations of the data mining models and presentations. According to another aspect of the invention, a method is provided wherein the step of creating presentations and data mining models is accomplished by a text editor, scripting tool, web development tool, or HTML editor. According to another aspect of the invention, a method is provided wherein the step of updating data mining models and presentations includes editing, modifying, deleting, replacing, suspending, changing, improving, and optimizing the data mining models and presentations.

According to another aspect of the invention, a data mining system for delivering presentations associated with data mining models is provided. This system includes a repository for storing data mining models and presentations; means for accessing, creating, updating, and importing data mining models, presentations, and the content and structure of the repository; means for selecting and preparing presentations for delivery to users; and, means for accessing the data mining system by users, for receiving first and second inputs from users, and for delivering presentations to users.

According to another aspect of the invention, a system is provided wherein data mining models include simple and complex scores, static and dynamic scores, rules, rules sets, and rules systems, and composites and combinations of such scores, rules, rules sets, and rules systems. According to another aspect of the invention, a system is provided wherein presentations include defined sets of content including data mining model references, scores, rules, attributes, activity suggestions, to do lists, pop-up windows, HTML segments, XML scripts, and sets of computer instructions. According to another aspect of the invention, a system is provided wherein presentations include simple presentations, range based presentations, and complex presentations.

According to another aspect of the invention, a system is provided wherein the repository is a database having a database schema and database management system. According to another aspect of the invention, a system is provided wherein the means for accessing, creating, updating, and importing data mining models, presentations, and the content and structure of the repository includes a data mining console system which incorporates all interfaces and related functionality required to configure the system by reference to a plurality of data sources, support the import and preparation of data extracted from these sources through templates and tools created for numerous industry, business, user and data types, and supports the manual or automated creation of data mining models by applying a plurality of data mining algorithms to these data sources.

According to another aspect of the invention, a system is provided wherein the means for selecting and preparing presentations for delivery to users includes an analytic decision engine system including model presentation services and scoring services modules. According to another aspect of the invention, a system is provided wherein the means for accessing the data mining system by users, for receiving first and second inputs from users, and for delivering presentations to users includes networks including the Internet.

According to another aspect of the invention, a system is provided wherein users include data mining management consoles integrated with or connected over networks. According to another aspect of the invention, a system is provided wherein users include applications running on a computer, system, or other appliance integrated with or connected over networks. According to another aspect of the invention, a system is provided wherein users include thin or thick clients in a client-server or browser-server environment integrated with or connected over networks.

According to another aspect of the invention, a system is provided wherein first inputs include attributes provided by or associated with users including identity and locator. According to another aspect of the invention, a system is provided wherein second inputs include updates, edits, modifications, deletions, replacements, suspensions, changes, improvements, and optimizations of data mining models and presentations. According to another aspect of the invention, a system is provided that includes reporting systems. According to another aspect of the invention, a system is provided

wherein the reporting systems include outcome monitoring systems and presentation usage monitoring systems. According to another aspect of the invention, a system is provided wherein the reporting systems are operable to produce second inputs.

5 According to another aspect of the invention, a system is provided that has stored therein data representing sequences of instructions which when executed cause the above-described method to be performed. The system generally has servers, memory devices, processing units, input devices, output devices, display devices, and users for storing and processing sequences of instructions and for enabling communications with users.

10

BRIEF DESCRIPTION OF THE DRAWINGS

The invention may best be understood by referring to the following description and accompanying drawings which illustrate the invention. In the drawings:

15 FIG. 1 is a block diagram illustrating a typical data mining system in accordance with the prior art;

FIG. 2 is a block diagram illustrating a data mining system in accordance with an embodiment of the invention;

FIG. 3 is a block diagram illustrating typical interfaces for the data mining system in accordance with an embodiment of the invention;

20 FIG. 4 is a block diagram illustrating exemplary interfaces to the model repository of the data mining system in accordance with an embodiment of the invention;

FIG. 5 (a) is a block diagram illustrating integration of a traditional client-server or “thick” desktop application with the analytic decision engine in accordance with an embodiment of the invention;

25 FIG. 5 (b) is a block diagram illustrating an interface suggestion applet in accordance with an embodiment of the invention;

FIG. 5 (c) is a screen capture of an interface suggestion applet in accordance with an embodiment of the invention;

FIG. 6 (a) is a block diagram illustrating integration via pop-up window of an application with the data mining system in accordance with an embodiment of the invention;

FIG. 6 (b) is a block diagram illustrating integration via pop-up window within the Siebel "Call Center" application in accordance with an embodiment of the invention;

FIG. 7 (a) is a flow chart of thin client server integration with a web server in accordance with an embodiment of the invention;

FIG. 7 (b) is a block diagram of frame based thin client server integration in accordance with an embodiment of the invention;

FIG. 8 is a block diagram illustrating a database schema for the model repository in accordance with an embodiment of the invention; and,

FIG. 9 is a block diagram illustrating exemplary reporting modules for the analytic decision engine in accordance with an embodiment of the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

In the following description, numerous specific details are set forth to provide a thorough understanding of the invention. However, it is understood that the invention may be practiced without these specific details. In other instances, well-known software, structures and techniques have not been described or shown in detail in order not to obscure the invention. In addition, the system and method may be implemented using a number of different programming languages (e.g. Java and Enterprise Java Beans) and for a variety of different platforms (e.g. UNIX and Linux). The following description illustrates the invention by reference to the Microsoft platform comprising the Windows operating system, SQL Server relational database server, and Internet Information Services web server, but other platforms, database

server, web servers, applications and programming tools may also be used. In the drawings, like numerals refer to like structures or processes.

5 The term “data mining system” (“DM system”) is used herein to refer to any machine for processing data including transmitting or receiving data, and including the computer systems, devices, and network arrangements described herein. The term “data mining algorithm”, also referred to as “knowledge discovery” algorithm, is used herein to refer to algorithms including, for example, logistic regression, linear regression, neural networks, Bayesian networks, decision trees, association rules, cluster algorithms (e.g. K-Means or Expectation Maximizing), that may be
10 incorporated into DM systems.

The term “data mining model” (“DM model”) is used herein to refer to a predictive model developed using one or more data mining algorithms. A data mining model may consist of a single score or set of scores, a single rule or set of rules, as well as composites and combinations of such scores and rules, comprising a complex rules
15 system.

The term “score” is used herein to refer to a probability of an event. For example, a score of 0.1 may reflect a 10% probability that a given customer will default on a loan).

20 The term “complex rules system” is used herein to refer to a list of potential outcomes and associated probabilities (e.g. a complex rules system for a “next purchase” transaction may predict the likelihood of several different outcomes based on a variety of different data inputs, including past actions, transactions, interactions, geodemographic data and other relevant data about both the transactor and other persons, with a confidence measure associated with each different possible outcome).

25 The term “data mining provider” is used herein to refer to an application server (e.g. ANGOSS KnowledgeSERVER™, Microsoft™ SQL Server Analysis Services, or an implementation of an application server under JSR 073, a Java Community standard) that may train data mining models (i.e. based on the analysis of data) and call existing data mining models to generate scores (i.e. based on a new data instance).

The term “static model” refers to a data mining model that calculates a score (for example, by looking up a pre-computed score stored in a table as a result of a batch process).

5 The term “dynamic model” is used herein to refer to a data mining model that calculates a score on a real-time basis. The score may be calculated using model inputs that are provided in the context of the person to person, person to system, or system to system interaction or looked up from a required data source while an interaction occurring within a digital ecosystem is in process.

10 The term “presentation” is used herein to refer to a defined set of content (e.g. a promotion, suggestion, script, document, a digitally reproduced image capable of being rendered in a computer interface, or a set of instructions capable of being processed by a computer system) together with a collection of data mining model references, attribute references, rules, and formulas that may define how to select and
15 prioritize that presentation and all other possible presentations based on an examination of any number of different simple or complex rules systems and attributes to determine the relevant reference score and attribute values for a particular person or entity. The presentation is the outcome or result of this automated decision process. The presentation may be represented by an HTML segment that is displayed through a computer interface or it may be represented by a set of instructions
20 communicated to one or more computer systems.

In general, the present invention provides a method and system for the creation, storage and management of data mining models (i.e. scores, rules and complex rules systems), including the development and deployment of data mining models, the association of data mining models in both static and dynamic modes with
25 presentations, and the continuous evaluation and optimization of these models and presentations for a plurality of computing environments. For example, using the invention described herein a data mining model may compute that a particular customer is both highly profitable and highly likely to defect. The present invention enables both the computation of these propensities, in either a systemic fashion
30 through the use of predefined configuration tools and templates to create such computation, or at the time of a person to person or person to system interaction involving that customer, and the serving of appropriate, defined presentations which

take into account these propensities and related business rules that have been established to address “high profitability”, “high risk” and “high profitability and high risk” customers (e.g. a directed offer to approve a waiver or reduction of interest rates or service charges) as an HTML segment where this HTML segment may provide for the display of the presentation and other possible presentations (i.e. including the suppression of all presentations), in priority order, as an HTML document. Alternatively, the presentation may comprise the execution of an automated process or the delivery of content to some other person or location to initiate an action or course of action. For example, a background process may monitor a web based customer filing process for new insurance claims and pass these claims through one or more data mining models (such as one or more of a loyalty (defection) model, a fraud model, and a profitability model). If the presentation computed is “INVESTIGATE”, for example, the process would route the claim to an investigator.

The invention, may be deployed within enterprise software applications, including, for example, the Siebel Systems, Inc.’s (“Siebel”) suite of “eBusiness Applications” (“eBusiness”). The Siebel eBusiness application suite is a leading “customer relationship management” (“CRM”) software application and is representative of standard CRM environments. The invention supports the creation of data mining models using tools and interfaces designed specifically by reference to the Siebel data schema, and the deployment of data mining models and associated presentations to Siebel eBusiness applications in several modes including applets working within the interface environment and “pop up” windows working within the Siebel eBusiness environment. Through integration with the Siebel eBusiness application infrastructure, the invention’s analytic decision engine treats all Siebel eBusiness interfaces (e.g. Siebel “eSales”, “eService”, “eCallCenter”, “ePersonalization”, and “ePricer”) as target environments for the delivery of presentations based on the deployment of data mining models. Integration options may include using Siebel applets working within the Siebel interfaces themselves, or with browser based messaging from the invention’s analytic decision engine to the target environment, or using the system itself as an integrated application within the Siebel eBusiness environment.

With the inventions when a data mining model is created, it may be integrated with an enterprise's systems where the knowledge provided by the model may be usefully applied through the delivery of appropriate content, suggestions, presentations or other presentations at the point of interaction (e.g. through the delivery of a notice about a special promotion for profitable customers to be given to a profitable customer visiting a web site, through the automated flagging and investigation of a transaction identified as suspicious immediately upon the reporting of the transaction, or through the timely delivery to a sales representative of a product or service sales suggestion for one of their client's based on the recent transaction history of that client).

Data Mining System Functional Components. Referring to FIG. 2, there is shown a block diagram of a data mining system **200** in accordance with an embodiment of the invention. The data mining system **200** includes the following: an analytic decision engine ("ADE") **210**; a model repository **220** for the storage of scores, rules, and rules sets resultant from multiple data mining tasks; a data mining management console **230**; and, means for importing into the repository **220** a wide variety of data mining models **240** and scores **250** generated from data mining processes involving diverse DM Tools (e.g. SAS Enterprise Miner, IBM Intelligent Miner, ANGOSS KnowledgeSTUDIO), RDBMS Systems (e.g. Microsoft SQL Server 2000), and DM Applications.

In addition, the data mining system **200** may include the following: a web or application server **260** that can serve as a target environment for the deployment of data mining models and presentations; customers **270** interacting in the digital ecosystem of an organization through the Internet who may be identified and targeted for the delivery of data mining models and associated presentations; representative thin client enterprise applications **280** regulating person-to-person, person-to-system, and system-to-system interactions (e.g. call center and sales force automation applications) which may be present in the digital ecosystem of an organization and which serve as presentation environments for the analytic decision engine **210**; representative client-server types of enterprise applications **290** regulating person-to-person, person-to-system and system-to-system interactions (e.g. "thick" client or "desk top" call center and sales force automation applications) which may be present

in the digital ecosystem of an organization and which serve as presentation environments for the analytic decision engine **210**; representative enterprise productivity applications **291** (e.g. marketing and campaign management applications), in either client-server or “thin client” forms, which may be present in the digital ecosystem of an organization and which serve as presentation environments for the analytic decision engine **210**; and, automated business process applications **292** (e.g. automated transaction and reporting systems such as claims systems) in either client-server or “thin client” forms, which may be present in the digital ecosystem of an organization and which serve as presentation environments for the analytic decision engine **210**.

The analytic decision engine **210** supports the import of models **240** and scores **250** for use in the analytic decision on a platform and data mining tool agnostic basis and may support a wide variety of DM Tools, RDBM Systems, and DM Applications.

In general, the analytic decision engine **210** includes means designed to accelerate the data mining process and provide a flexible, extensible, and highly scalable environment for deploying and managing a number of different data mining models and associated presentations. It may support, for example, the following: (a) the importing of data mining models **240** and scores **250** together with conversion utilities for converting proprietary data mining models (e.g. SAS models) into XML documents to support applications of the analytic decision engine **210** on a vendor, platform, and system neutral basis; (b) a model repository **220** for the storage of these data mining models **240**, as well as for optimization of these models and supporting other features of the system; (c) a data mining management console **230** which provides means to enable a human user to manage, evaluate, and optimize a plurality of data mining models **240** operating in one or more digital ecosystems; and, (d) an application server **260** which supports the deployment of data mining models to users and interfaces connected to the application server directly by wired or wireless connections **290**, **291**, **292** or through an application, interface, or appliance connected to the network by wired or wireless connections **270**, **280**.

In more detail, the data mining system **200** and analytic decision engine **210** allow for the storage of data mining models, consisting of simple and complex rules systems, rules, rules sets and rules systems, as well as composites and combinations of such

scores, rules, rules sets and rules systems (all of the foregoing being data mining models), through computer-based graphical user interfaces, network connections, and means for:

- 5 a) Directly creating data mining models for storage in, or importing data mining models developed using a variety of data mining tools into, a common repository (i.e. the model repository **220**), in a file format or database schema, and accessible concurrently to one or more computers in a local area network, wide area network, or Internet network **260, 270, 280, 290, 291, 292**;
- b) Populating the model repository **220**;
- 10 c) Exposing the schema to permit two way access via the network by human users or executable programming code in the form of software applications or interfaces running on computers, systems, and other appliances integrated with or connected over the network **260, 270, 280, 290, 291, 292** to the model repository **220**;
- 15 d) Managing applications across the network **260, 270, 280, 290, 291, 292** with the model repository **220** for model creation, interrogating the model repository **220**, and returning processing results;
- e) Optimizing the continuous accessing, testing and modification of the data mining models **240, 250** contained in the model repository **220** as a result of
20 user inputs to, and or computer managed interactions captured by, applications **260, 270, 280, 290, 291, 292**; and,
- f) Providing user interfaces for the above.

Scoring Services and Presentation Services. Referring to FIG. 3, there is shown a block diagram illustrating typical interfaces **300** for the data mining system **200** in
25 accordance with an embodiment of the invention. The primary modules within the ADE **210** include presentation services **211** and scoring services **212**. The ADE **210**, presentation services **211**, and scoring services **212** integrate with the typical interfaces **300** which may include the following: a browser **310**, desktop applications **320**, both of which may communicate with Windows 2000/IIS (internet information

server) **340** via HTTP **330**; COM+ (component object model extension) **350** links between Windows 2000/IIS **340** and the analytic decision engine **210**; OLEDB-DM (application program interface) links between the analytic decision engine **210** and OLEDB-DM servers **390**; and, HTML, XML, COM (component object model),
5 DCOM (distributed component object model), or SOAP (simple object access protocol) **360** links between the analytic decision engine **210** and other servers and clients **260**.

Referring to FIG. 8, there is shown a block diagram of a database schema **800** for the model repository **220** in accordance with an embodiment of the invention. The
10 database schema **800** is referred to as the "ASYS" database in the following. Referring to FIGS. 2, 3, and 8, the analytic decision engine **210** integrates with an organization's information technology environment via the presentation services ("MPS") **211** and scoring services ("SS") **212** modules.

The MPS module **211** includes the following functionality:

- 15 • Generation of HTML to display in a presentation environment of a client-server or browser server based application such as on a web site or to a customer service representative (i.e. in the manner which will be described below with reference to FIGURES 4 through 7).
- 20 • Monitoring the use of data mining models by providing an audit file (i.e. as part of the repository **220** a representative schema of which is illustrated in FIG. 8) which records relevant data with respect to the use of simple and complex models with related means to define parameters for such recording actions (e.g. frequency) by reference to specific models and other measures. This functionality enables a user to manage a wide range of simple and
25 complex models operating in a variety of different systems and to assess the performance of these models.
- 30 • The incorporation of appropriate content into the HTML being presented that, in conjunction with server modules, records the outcome of the running of the models and presentations in the repository **220**, as illustrated in FIG. 8, for use in training new models, in optimizing existing models, and in modifying or replacing presentations associated with existing models.

- The MPS may also return the presentation ID list as opposed to the HTML which conveys those presentations. This capability allows for additional features described below.
- The ability to benchmark performance of MPS 211 and SS 212 by applying a long list of customers in sequence to a presentation definition. This may be accomplished by running a data batch processing routine that performs the same operations done in a real time mode. Performance statistics may be gathered and analyzed by running the presentation definition in this batch mode to estimate performance and hardware requirements for a real time system.
- The ability to generate overview reports that show the overall distribution of the presentations. A presentation editor may allow users to create a complex series of rules and formulas that determine which presentation to give to a particular individual. As such, it is easy to loose site of the over all effect of these. By applying the presentations to a complete list or statistically representative sample of customers and tabulating the resultant presentations, a user may gain an overall view of the presentations.
- The ability to export a list of presentations that may be imported by a marketing automation application. This allows an organization to have a unified presentation strategy across all marketing channels.
- In order to achieve high scalability, an instantiated MPS object may be heap-lock free (i.e. runtime read only), multithreaded, and initialization stateless. Instantiating and initializing an MPS object is time consuming. It may digest all the presentation rules and formulae, pre-compile any SQL data acquisition or scoring commands, and load the presentation HTML into memory. As such, it should not be initialized during each score request. Instead, the object should be instantiated once when it is first used or at another appropriate time. This means that the object may exist on the server for an extended period of time (e.g. stored in the Windows Running Object Table or ASP Application object) and many different users (i.e. execution threads) may be simultaneously using the MPS object. In order for this not to cause

performance bottlenecks, the MPS object should not require many (any) mutex locks required when writing to heap memory.

- 5 Because MPS is designed for operational environments and live objects may be retained in memory, updates to presentation definitions may implement special handling. In order to handle this situation, presentation objects have a version number (refer to the “ASYS_Projects” table 810 contained in FIG. 8) and the moniker (name) of the object has the ID and version number. When a new version of the object is saved, the version number is incremented. The next time the presentation object is requested the new most up-to-date one will be instantiated and initialized. If any old objects are still in use, they will run as normal.
- 10
- 15
- 20
- 25
- 30
- When a calling module (e.g. client browser or another server) calls MPS to request the results of a presentation, a “Model Locator” may be used to refer to the presentation. This locator in turn links to one or more presentations. This allows a presentation to be changed without having to modify any clients. For example, all clients could use the URL query string “PresentationLocator=StdCC&EntityID=12345”. In this case, “StdCC” is the ID of the locator. This locator ID is looked up to discover the actual presentation. In order to have clean and unbiased data on the benefits of one presentation versus another, the locator ID may actually link to more than one presentation. One of the presentations may be a test group. Some exposures will be given to the test presentation on a random or pre-selected customer list basis. This allows for valid comparisons between two or more presentations.
 - The client may also pass data to the ADE 210 for use by dynamic models. The query string “PresentationLocator=StdCC&EntityID=12355&Age=25&Sex=M” is an example of this. In this example, if any of the dynamic models require inputs “Age” or “Sex” then the passed values will be used, even if a data lookup can also fulfill the data request. Furthermore, the query string, “PresentationLocator=StdCC&Age=25&Sex=M” does not include an entity ID. In this case, all the required data must be provided and all the models

must be dynamic. This feature allows for “anonymous CRM” (i.e. client relationship management).

The SS module 211 includes the following functionality:

- 5 • The ability to deal with any number of different data mining scoring engines and model types. This is accomplished by a portability layer which will be referred to as the real time scorer (“RTS”). RTS may be initialized with a list of model ID’s. Each model may be associated with a particular data mining provider (refer to the “ASYS_MODEL” and “ASYS_MODEL_SOURCE” tables 820, 830 contained in FIG. 8). RTS may in turn call the actual Data
10 mining providers to calculate the scores.
- To optimize performance, SS may have similar characteristics to the MPS module in that an SS object may be thread safe, initialize stateless, and read only to heap memory.
- The SS may expose an identical interface for both static models and dynamic
15 models. This allows dynamic and static models to be easily interchanged.
- A series of data fulfillment routines may be associated with a dynamic model (refer to the “ASYS_ModelInputLookup” table 870 in FIG. 8). The SS may be initialized with 1 or more models and their associated data fulfillment data. The data fulfillment requirements may be optimized taking into account
20 similarities between models. For example, if four models require an “Age” variable, it need only be looked up once.
- A common data acquisition interface may be used to fulfill model data requirements. An example of such an interface is as follows:

```
25           interface IGenericLookup : IDispatch
          {
              [id(1), helpstring("method Initialize")]
HRESULT Initialize([in] long lProjModelID, [out, retval]
VARIANT_BOOL *ret);
              [id(2), helpstring("method Prepare")] HRESULT
30       Prepare([in] BSTR sConnStr, [in] BSTR sUserID, [in] BSTR
sPassword, [out, retval] VARIANT_BOOL *ret);
```

```

[id(3), helpstring("method GetData"),] HRESULT
GetData([in] BSTR sEntityID, [out, retval] VARIANT *ret);
[id(4), helpstring("method GetCommandText")]
HRESULT GetCommandText([out, retval] VARIANT *ret);
};

```

- A prepare method allows any time-expensive operations (e.g. compiling SQL or integrating data files for lookup information) to be done during an initialization phase.
- By using such an interface, new data lookup techniques may be added.
- Three exemplary interfaces are as follows:

1. Data Lookups

- a) The connection string, table, field name (or expression), ID and further WHERE clause info are stored in the definition (refer to the “ASYS_ModelInputLookup” table **840** in FIG. 8).
- b) A SQL statement is generated from this information.
- c) The effect of different connection strings is “joining” information from a variety of different sources.
- d) If more than one model requires the same information it is consolidated.

2. SQL

- a) Users may also enter a SQL statement to calculate any data requirements (refer to the “ASYS_ModelInputSQL” table **850** in FIG. 8)

3. Stored Procedures

- a) This allows a database stored procedure to be specified and executed as part of the data fulfillment process (refer to the “ASYS_ModelInputSP” table **860** in FIG. 8)

30 *Implementation of the Data Mining System Using the MPS and SS Modules.* Referring to FIG. 4, there is shown a block diagram of exemplary interfaces **400** to the model repository **220** of the data mining system **200** in accordance with an embodiment of

the invention. The exemplary interfaces **400** include system management interfaces, integration with authoring tools for presentation creation, and illustrative interfaces integrated with the MPS **211** and SS **212** modules. These interfaces **400** include the following:

- 5 • interfaces **410** and data processing routines enabling the creation, management and optimization of intelligent business rules systems to support customer service requests and business process automation (e.g. credit limit increases, fee and rate adjustment requests, and approval and pricing decisions) through the integration and deployment of data mining models and associated
10 presentations at relevant points of interaction;
- interfaces **420** and data processing routines enabling relationship managers, investigators, and knowledge workers to access, via their browser, presentations of relevance to their functional areas of business activity (i.e. their client base and targeted prospects) to support their daily activity and
15 improve their overall productivity;
- interfaces **430** and data processing routines enabling the deployment of data mining models in both static and dynamic mode to customer service representative interfaces (e.g. contact management systems and call center systems) and customer self-service interfaces (e.g. the web and Internet or
20 Intranet based kiosks);
- interfaces **430** and data processing routines enabling the creation and storage in the Data Mining Repository of HTML documents (such as sales scripts, marketing promotions, forms, and other useful guides and tools) created for association with data mining using industry standard HTML authoring tools
25 scores to support targeted interactions;
- presentation services **440**;
- interfaces **450** and data processing routines enabling the capture, storage and analysis of feedback gathered from interactions triggered by the calling of data mining models and associated presentations from the Data Mining Repository

to support optimization of these interactions and the business processes supporting them;

- interfaces **450** and data processing routines enabling the monitoring and creation of reports to assess the performance of data mining models and defined presentations in these target environments; and,
- model deployment services **460**.

In general, the exemplary interfaces **400** include the following: means for the importing of data mining models and scores (i.e. derived from DM Tool RDBMS System or DM Application) **240, 250**; a model repository **220**; analyst interfaces **410** to manage simple and complex models and scores in the model repository; knowledge worker productivity interfaces **420** to enable the association of simple and complex models and scores, as well as attributes, with presentations using HTML editors; model deployment and optimization interfaces and services **430** to enable the deployment of batch and dynamic models to any appliance, interface, application, database, or system, the tracking and analysis of outcomes, and the optimization of data mining models and scores contained in the model repository based on these outcomes; presentation interfaces and services **440** to enable the deployment of presentations associated with static and dynamic data mining models, the tracking and analysis of outcomes, and the optimization of data mining models and scores contained in the model repository based on these outcomes; decisioning interfaces and services **450** to enable the automation of actions and interactions supporting business processes by automating the application of data mining models to these actions and interactions, the tracking and analysis of outcomes, and the optimization of data mining models and scores contained in the model repository based on these outcomes; and, static and dynamic model deployment in batch and real time modes is enabled to support any action or interaction, across the Internet, through any network, to any appliance, application, or interface **460**.

The MPS **211** and SS **212** modules support the efficient integration of data mining models resulting from the analysis of a wide variety of different data sources into a common environment for the management of these models and for the association of

a wide range of different presentations linked to the data mining models in an organization's operating environment.

As referred to above, a "presentation" may consist of a targeted message, promotion, or other relevant content which supports a person-to-person, person-to-system, or system-to system interaction (e.g. a promotion to a customer or prospect which meets the criteria associated with the underlying data mining scores defined for that presentation) or it may consist of content served internally within the organization (e.g. a list of claims or recent transactions which warrant investigation routed by MPS to internal claims review or other investigatorial personnel). MPS 211 provides this functionality as follows:

- MPS 211 includes the ability to manage the display of presentations through:

- XML documents that embody the rules and calculations to determine which presentations (and which HTML) to display in relation to a particular customer or event;

- Management of a diverse range of presentations contained in the repository, with each possible presentation having:

- An ID;
- An XHTML segment for displaying the presentation; and,
- An SQL statement for suppressing the presentation in circumstances where the user wishes to not run a presentation otherwise suggested by the results of a data mining model (e.g. to reflect recent events not reflected in the model such as an offer to purchase a certificate of deposit if a customer has just purchased a certificate of deposit);

- Formatting information for the score. For example, if it is a simple score, a user may wish to display the score as a percentage or as a number from 1 of 5. Complex rules systems require enough formatting information to produce an HTML table from the complex rules system;

- Management of the process through which feedback is obtained, logged, and analyzed in response to the use of presentations, including:
 - Whether to implement a feedback procedure in conjunction with the Presentation;
 - The manner of implementing the feedback procedure; and,
 - HTML to support the selected feedback mechanism;

5

- MPS 211 also enables a user to incorporate a number of different presentation options or alternatives into a presentation and to define the rules for selecting from among these alternative presentation options, including:

10

- Ordering and filtering presentation types (e.g. on a top n basis) to permit optimization based on incorporation of multiple data mining models through the calculation of combined scores and then selecting the presentation defined for the best combined score or combination of scores;

15

- MPS 211 also provides for combination of data mining models into combined scores, for example, for use in situations where an organization wishes to assess the best possible outcome (e.g. a best offer to a customer) in circumstances where multiple variables and multiple data mining models influence the determination of this best offer and in other areas where rules based optimization of presentations is desired, through permitting the computation of combined scores, new scores, and other presentation constants by calculating these values from the data mining models stored in the model repository.

20

25

In this manner, MPS 211 provides for the following:

- Simple presentations (i.e. presentations with one presentation option only, based on one data mining model or combination of data mining models). For example, consider the situation where a telecommunications customer with high profitability is showing a high likelihood of defection and is interacting with a call center representative. A presentation may be delivered via HTML

30

to the call center representative at the time of interaction suggesting a lower cost or longer term program to address the risk of defection;

- Ranges based presentations (i.e. presentations with several presentation options, each of which is based on a range of scores contained in one data mining model or a combination of data mining models); and,
- Complex presentations. That is, presentations with several presentation options, based on a range of scores contained in a data mining model or data mining models together with any number of defined attributes (e.g. data values, which can be text, numeric, dates or other simple data types, such as age, gender, product ownership, etc.) that are looked up in real time while actions, transactions and interactions are occurring to assist in selecting the presentation. As part of MPS, since a complex presentation has many permutations and combinations of possible presentations associated with it, user interface components are provided representing the rules based hierarchy for presentation delivery to assist in rules formulation and rules editing. For example, a decision tree predictive model may be used to represent a rule based tree structure, and a decision tree editor may be used to formulate the rules. Each terminal node of the tree leads to a specific presentation. This approach allows the data mining provider to execute the rules to determine the correct presentation.

MPS 211 also provides for the ordering of different presentations (e.g. through tagging and sorting presentations based on model scores as “Beginning”, “Sorted”, or “End”.) As a result, the analytic decision engine 210 provides for the following:

- the application, via the scoring services API 212 within the analytic decision engine 210, of predictive models to transactional systems, for example through the reading and analysis of transactions from a message queue service, such as IBM’s MQSeries or Microsoft’s Message Queuing Service (MS MQ), enabling the identification of predicted outcomes (e.g. fraud, defection, best bid or other valuable knowledge relative to business decision-making);
- the exposing of these scores for interrogation by MPS 211; and,

- the resultant serving of instructions to MPS 211 which proceed to serve appropriate HTML in the form of messages, lists, suggestions or any other defined HTML content wished to be presented to support a transaction, interaction or other action based on the scores defined and determined within SS 212 enabling the triggering of appropriate actions (e.g. routing transactions determined to have a high fraud score via SS 212 to investigators through MPS 211).

In addition, MPS 211 module types and SS 212 module types may be accessed with HTTP, COM, or SOAP interfaces. Moreover, the data mining management console 230 included with MPS 211 may provide the wizards and dialogues to enable users to direct the functionality provided by MPS and SS module types. Examples include: the importing of static or dynamic models; the interrogation of the model repository to assess the behavior and performance of such models in production environments within the digital ecosystem; the generation of reports for distribution in the form of HTML with respect to model performance; the results obtained from serving presentations based on the scoring services and presentation services modules; etc.

By enabling MPS 211 and SS 212 to be accessed with HTTP, COM, SOAP, and XML as a representative integration platform, the system allows the analytic decision engine 210 to service a wide variety of clients and servers. The analytic decision engine 210 is based on industry standards (e.g. Microsoft and Internet technologies, including Windows 2000 and COM+ MTS) which make the analytic decision engine 210 highly scalable. Access to the repository 220 may be accomplished via a programmatic interface that enforces accessibility rules.

In addition, the MPS 211 and SS 212 modules provide means through which relationship management, case management, exceptions based analysis, and other business processes may be managed through the incorporation of relevant data mining models and delivery of associated presentations.

For example, a “relationship manager” view may consist of the following:

- A parameterized SQL query to build a list of cases (e.g. customers, claims, etc.) for each interested party;

- A link to the presentation to use;
 - Display formatting parameters such as the following:
 - Does user want to display presentations in the table;
 - Does user want to precompute the presentations to maximize performance (refer to the “ASYS_RelManPrecomputes” table 880 in FIG. 8);
 - HTML page expiry (determines when HTML is regenerated);
 - Does user want to pre-compute and store the resultant HTML; and,
 - Ten most likely to defect, etc.
- 10 Again referring to FIG. 2, the data mining management console 230 provides user interface components enabling data mining modelers, business analysts, and others to manage the data mining system 200. These user interface components, which may be implemented in the form of a hierarchical management console providing appropriate authorization and view screens for different individuals involved in the business
- 15 process optimization, include the following features:
- Wizards and dialogs to enable a user to define views and edit queries;
 - Batch process routines to enable the pre-computation of presentation lists;
 - The ability to preview other views in the data mining system (e.g. relationship manager and case presentation views); and,
 - 20 • The ability to administer file and referral URL locators for the management of presentations and presentations.

The analytic decision engine 210 enables the efficient management of data mining systems in a distributed computing environment. The analytic decision engine 210

25 includes the following functionality:

1. Provides a reusable way of displaying tabular data on the browser;

2. Avoids frequent refreshes from the web server for small to medium sized reports thus increasing performance; and,
3. Provides these reports as XML data so that these reports are available long after the transactional data has been purged

5

These functions make use of industry standards (e.g. XML) to interoperate with generally available browser solutions (e.g. Internet Explorer). Documents incorporated in the data mining system may be formatted based on XML standards and served in this format. Optimizations and transformations may be implemented with these standards. Since transformations are performed in the browser, performance is improved.

10

15

The analytic decision engine **210** also incorporates reporting functions to enable users to assess and monitor the performance of data mining models in operational environments. The primary components of the data mining management console **230** that provide reporting functions are as follows:

1. Stored procedures generate reports and return report data as ADO recordsets;
2. ASP pages generate the XML from these recordsets and store them in the data mining repository **220**; and,
3. Generic client side viewer ASP pages and generic client side viewer XSL scripts are used to enable the configuration of reporting systems and the display of reports on any browser.

20

25

The data mining management console **220** also includes the following functionality:

- The ability to provide page displays on the browser in a customizable manner with support for standard “skins” as frames for reports or to fully customize the “look and feel” of page displays at or following initial installation and configuration; and,

30

- Optimizations to support processes locally on the browser, rather than through server hits, when interacting with page views (e.g. by providing for sorting of records on a browser when a field header is clicked, rather than via a server side process).

5

Referring to FIG. 9, there is shown a block diagram of exemplary reporting modules **900** for the analytic decision engine **210** in accordance with an embodiment of the invention. Reports provided through the data mining management console **230** may be based on information contained in the “ASYS_Feedback” table **890** forming part of the model repository **220**. These reports may assess the performance of data mining models and presentations in the data mining system **200**. They may also provide measures and details of when data mining models have been used and how they have performed. They may include user functions relating to statistics, measures, and sorting. In FIG. 9, an ADO recordset **910** is converted to XML **920** which is stored in the repository **220** (e.g. ASYS_ModelFeedback **890** and ASYS_ModelUsage **891**). This information may then be delivered to a web server **940** and viewed using a browser **950**.

For example, one reporting function provided by the analytic decision engine **210** is a distribution report, which provides statistics on the types and performance of data mining models stored in the repository **220**. Fields included in this report may include the following:

- Model ID;
- Model Description;
- Average Score. This is the average score for the model. Data from multiple reports of the same kind can be used to see if the model performance is degrading over a period of time or improving;
- Standard Deviation Score;
- Cardinality of the Score. This would be the number of distinct scores for the model; and,
- Number of missing scores

The generation of this report may be provided through stored procedures (e.g. ASYS_sp_GenModelScoreDistReport).

Integration of Analytic Decision Engine with Desktop Clients. The integration of data mining models and presentations with user interfaces and other presentation environments contained in traditional client-server (i.e. "thick client") and browser-server (i.e. "thin client") applications, running on a variety of different operating systems in a plurality of different personal computer, PDA, and other appliances, will now be described with reference to FIGS. 5 (a) through 7 (c).

Referring to FIG. 5 (a), there is shown a block diagram illustrating the integration of a traditional client-server or "thick" desktop application with the analytic decision engine **210** in accordance with an embodiment of the invention. This integration may be accomplished through the use of a browser Active X control embedded in the application which is configured to determine (i.e. through the data mining system **200**) and display a presentation. Referring to FIGS. 2, 3, and 5(a), desktop applications **320** running on personal computers may gain access to the analytic decision engine **210** using a Browser ActiveX control that is included as a standard component of Microsoft Internet Explorer (i.e. version 4 or 5) ("IE"). For non-IE environments, the analytic decision engine **210** may be configured to support the Mozilla browser which is an open source implementation of the Netscape browser. These controls may be distributed to users who do not have or want IE. Once the Browser ActiveX control is placed in a desktop application **320**, a user may communicate with the presentation services **211** by navigating to the correct URL (uniform resource locator). As such, the same solution may service both thin and desktop applications.

Moreover, a user may employ HTTP **330** to communicate with the analytic decision engine **210**. The following are example URLs for static models, dynamic models, and mixed static and dynamic models, respectively:

`http://ScoreServer/Score.asp?EntityID=123&PresentationID=MyPres`

`http://ScoreServer/Score.asp?Age=12&Sex=M&PresentationID=MyPres`

`http://ScoreServer/Score.asp?Age=12&Sex=M&EntityID=123&PresentationID=MyPres`

Referring to FIG. 5 (b), there is shown a block diagram illustrating an interface suggestion applet **500b** in accordance with an embodiment of the invention. This is an example of desktop application integration, as illustrated in FIG. 5 (a), with reference to the Siebel "Call Center" application which is a traditional client-server application developed by a third party software developer.

Referring to FIG. 5 (c), there is shown a screen capture of an interface suggestion applet **500c** in accordance with an embodiment of the invention. This is an example drawn from the Siebel Call Center application as illustrated in FIG. 5 (b). The screen capture **500c** illustrates the analytic decision engine's **210** suggestion applet **500b** running in a standard "customer relationship management" software application, namely, the Siebel "Call Center" environment. In this example, a customer service representative user ("CSR") is provided with dynamic suggestions with respect to the best offer to make to the customer, who the CSR may be currently interacting with, based on the use of the analytic decision engine **210** to deliver data mining model outputs and associated presentations (in this case a customer offer) to the CSR. The analytic decision engine **210** supports this form of integration with a wide variety of different enterprise applications developed using industry standard development tools and programming languages (e.g. Microsoft Visual C++, Visual STUDIO, Visual Basic, Java, etc.).

The analytic decision engine **210** may be integrated with software applications in a variety of ways, taking into account the method of development and deployment of the broad range of different client-server ("thick client") and web-server ("thin client") software solutions. Two of these methods, which may be illustrated with reference to the Siebel enterprise application platform, are via suggestion applet or pop-up window. With respect to suggestion applets, and referring to FIGS. 5 (b) and 5 (c), a user may employ Siebel tools to create suggestion applet **500b** calls to the analytic decision engine **210**. This suggestion applet **500b** may contain the user's choice of browser ActiveX control and may call the presentation services **211**, as described above, with additional template code. Using this approach, suggestions appear as part of the Siebel application. This is advantageous when data mining forms part of development specifications or when major changes are being made to the Siebel application.

Referring to FIG. 6 (a), there is shown a block diagram of the integration via pop-up window **600a** of an application with the data mining system **200** in accordance with an embodiment of the invention. Here, the application integrates with the data mining system **200** through the use of a pop-up window **600a** working within the application to display a presentation. The pop-up window **600a** may be automatically invoked either by the user (e.g. by moving a mouse pointer to a designated icon in the application), on an automated basis without user action or intervention, or on an automated basis in response to a specific action or constraint. Referring to FIG. 6 (b), there is shown a block diagram of the integration via pop-up window **600b** within the Siebel "Call Center" application in accordance with an embodiment of the invention. Here, the user's position within the application is monitored by the data mining system **200**. If the user's current position is bound to a presentation (i.e. determined by interrogating the data mining system), then the pop-up window containing the presentation is automatically invoked. In addition, the data mining system enables changes in bindings for presentations without coding changes through the updating of database tables in the data mining system.

In other words, and referring to FIGS. 6 (a) and 6(b), the suggestion window may be a pop-up window **600a**. In this case, the application monitor **610b** shown in FIG. 6 (b) polls the Siebel application every second or two. When the user enters an area that warrants a suggestion, a window pops up. Where and when the window is popped up is managed via the data mining management console **230** as described above. This approach has the advantage that new suggestions may be added to various locations without Siebel repository changes (i.e. except, of course, for a simple initial change). This embodiment is advantageous for adding data mining capabilities to existing applications and for deploying predictive intelligence out to business users "on their terms" and through the interfaces that they are comfortable with. The embodiment has the following additional advantages: predictive models and related score defined presentations may be deployed across all Siebel applications in a common and consistent manner; there is no disruption to the Siebel data model or operational environment; and, outcomes may be moved directly to the analytic data mart for analysis, iteration, and optimization. The embodiment also includes the following advantageous features: platform independent; database independent; data mining and modeling tool independent; easy to implement and rapidly deploy; easy to manage

and monitor at both business and technical levels; and, easy to implement from a training perspective because front office customer facing personnel and customer touching CSRs are presented (i.e. through their browsers or other appliances) with content and outcomes only. Moreover, the analytic decision engine **210** remains CRM platform and CRM interface agnostic. The analytic decision engine **210** may support deployment of predictive intelligence to legacy environments, other CRM application platforms, and point solutions designed to fulfill defined requirements of an organization's CRM business and technology plans not addressed by Siebel or other similar third party CRM application vendors.

Referring to FIG. 7 (a), there is shown a flow chart **700a** of thin client server integration with the web server **260** in accordance with an embodiment of the invention. Here, the integration enables the delivery of HTML defined by the data mining system **200** (e.g. a personalized web page). At step **710a**, a page is requested. At step **720a**, a web server **260** receives the request and assembles the content. At step **730a**, the web server **260** calls the ADE server **210** and passes inputs including customer ID and presentation locator. At step **740a**, the ADE **210** computes the HTML for the presentation. At step **750a**, the web server **260** embeds the HTML in the presentation and returns the page. At step **760a**, the process ends or begins again.

Referring to FIG. 7 (b), there is shown a block diagram of frame based thin client server integration **700b** in accordance with an embodiment of the invention. Here, the method illustrated in the flow chart **700a** of FIG. 7 (a) is implemented by reference to communication between the browser (i.e. client) and the web server hosting the data mining system **200**.

Data Mining Repository and Data Mining Management Console. Again referring to FIG. 8, the data mining repository **220** stores data relating to the analytic decision engine's **210** scoring services **212** and presentation **211** modules. This repository **220** includes the following:

- Simple and complex "Scores"; "Presentations" (or instructions for locating such Presentations), including related presentation rules, which may be any number of different HTML based forms of documentation, links, instructions, images or other content determined to be or desired to be associated with

- 5 scores contained in the model repository; “Constraints” being any one or more user defined or system defined rules, rules system or constraints which supplement data mining defined predictive model scores, rules and rules sets with other user defined or system defined rules and rules sets to manipulate the manner in which Presentations based on Scores are used or not used; and, “Outcomes”, which may be numeric or text based and which are associated with the serving of one or more Presentations through time as described above for purposes of assessing and monitoring the performance of the system;
- 10
- Conversion utilities for converting proprietary data mining models (e.g. SAS models) operating in production environments into XML documents to support applications consuming predictive models on a vendor, platform, and system neutral basis;
- 15
- Support for the use of any combination of static or dynamic models or scores, together with composites thereof and other user defined constraints and rules as inputs in the analytic decision engine rules engine;
- 20
- Support for the integration of scores with user defined presentations to be served in connection with any person to person, person to system, or system to system interaction in the form of HTML, together with related optimization capabilities enabling the monitoring of results from the deployment of such models, scores and/or presentations to enable the continuous improvement of the capabilities of the system; and,
- 25
- An architecture suitable for a wide range of computing platforms and based on highly scalable industry standard technologies to enable the convenient implementation of the invention (i.e. availability of COM and SOAP interfaces for cross-platform implementation; use of COM+ MTS, HTTP, and XML for high scalability; client side JavaScript libraries to call the scoring server for dynamic adjustment to HTML forms and content based on model results).

Although the invention has been described with reference to certain specific embodiments, various modifications thereof will be apparent to those skilled in the art without departing from the spirit and scope of the invention as outlined in the claims appended hereto.

THE EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:

1. A method for delivering presentations associated with data mining models comprising the steps of:
 - a) storing said data mining models and said presentations in a repository in a data mining system;
 - b) associating ones of said data mining models with ones of said presentations through references within ones of said presentations;
 - c) receiving a request from a user for a presentation;
 - d) processing said request to determine an appropriate presentation corresponding to said request; said request including a customer identification; said request including an event identification; and,
 - e) delivering said appropriate presentation to said user.
2. The method of claim 1 and further comprising the steps of:
 - a) creating a database schema for said repository;
 - b) creating said data mining models and importing said data mining models into said data mining system; and,
 - c) creating said presentations and importing said presentations into said data mining system.
3. The method of claim 1 and further comprising the steps of:
 - a) receiving inputs from said user and from reporting systems included in said data mining system;

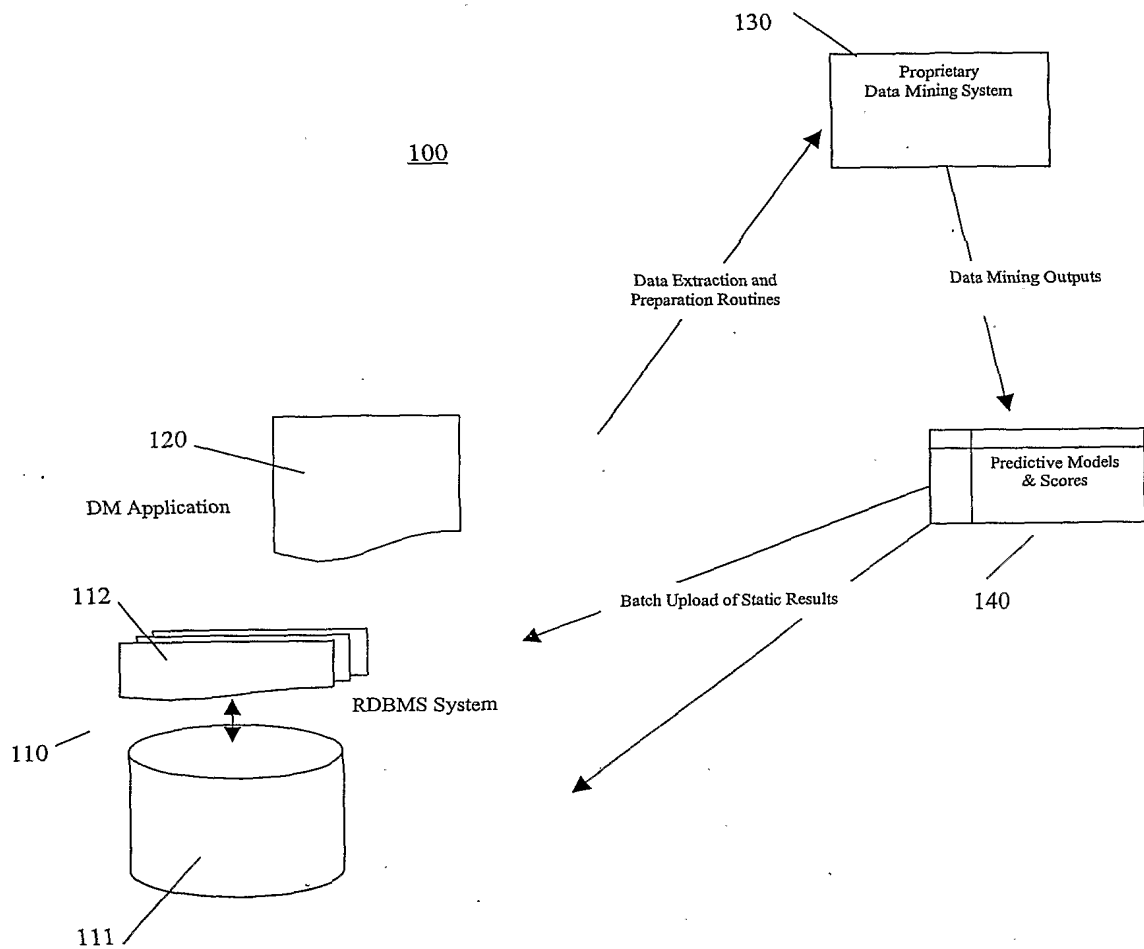
- b) updating ones of said data mining models and ones of said presentations based on said inputs to produce ones of updated data mining models and ones of updated presentations; and,
 - c) storing ones of said inputs, ones of said updated data mining models, and ones of said updated presentations in said repository.
4. The method of claim 1 and further comprising the steps of:
- a) defining rules under which said data mining models are associated with said presentations; and,
 - b) defining rules under which ones of said presentations are selected and prepared for delivery.
5. The method of claim 1 wherein said data mining models include simple and complex scores, static and dynamic scores, rules, rules sets, and rules systems, and composites and combinations of such scores, rules, rules sets, and rules systems.
6. The method of claim 1 wherein said presentations include defined sets of content including data mining model references, scores, rules, attributes, activity suggestions, to do lists, pop-up windows, HTML segments, XML scripts, and sets of computer instructions.
7. The method of claim 1 wherein said presentations include simple presentations, range based presentations, and complex presentations.
8. The method of claims 1, 2, 3, or 4 wherein said steps are accomplished over network means including Internet means.
9. The method of claim 1 wherein said user includes data mining management consoles integrated with or connected over said network.

10. The method of claim 1 wherein said user includes applications running on a computer, system, or other appliance integrated with or connected over said network.
11. The method of claim 1 wherein said user includes thin or thick clients in a client-server or browser-server environment integrated with or connected over said network.
12. The method of claim 3 wherein said inputs include attributes provided by or associated with said user including identity and locator.
13. The method of claim 2 wherein said step of creating said presentations and said data mining models is accomplished by means including text editor, scripting tool, web development tool, and HTML editor.
14. The method of claim 3 wherein said reporting systems within said data mining system include outcome monitoring systems and presentation usage monitoring systems.
15. The method of claim 3 wherein said inputs include updates, edits, modifications, deletions, replacements, suspensions, changes, improvements, and optimizations of said data mining models and said presentations.
16. The method of claim 3 wherein said step of updating said data mining models and said presentations includes editing, modifying, deleting, replacing, suspending, changing, improving, and optimizing said data mining models and said presentations.
17. A data mining system for delivering presentations associated with data mining models, said data mining system comprising:
 - a) a repository for storing said data mining models and said presentations;

- b) means for accessing, creating, updating, and importing said data mining models, said presentations, and the content and structure of said repository;
 - c) means for selecting and preparing ones of said presentations for delivery to users; and,
 - d) means for accessing said data mining system by said users, for receiving first and second inputs from said users, and for delivering said presentations to said users.
18. The system of claim 17 wherein said data mining models include simple and complex scores, static and dynamic scores, rules, rules sets, and rules systems, and composites and combinations of such scores, rules, rules sets, and rules systems.
19. The system of claim 17 wherein said presentations include defined sets of content including data mining model references, scores, rules, attributes, activity suggestions, to do lists, pop-up windows, HTML segments, XML scripts, and sets of computer instructions.
20. The system of claim 17 wherein said presentations include simple presentations, range based presentations, and complex presentations.
21. The system of claim 17 wherein said repository is a database having a database schema and database management system.
22. The system of claim 17 wherein said means for accessing, creating, updating, and importing said data mining models, said presentations, and the content and structure of said repository includes a data mining console system.
23. The system of claim 17 wherein said means for creating said presentations and said data mining models includes text editors, scripting tools, web development tools, and HTML editors.

24. The system of claim 17 wherein said means for selecting and preparing ones of said presentations for delivery to users includes an analytic decision engine system including model presentation services and scoring services modules.
25. The system of claim 17 wherein said means for accessing said data mining system by said users, for receiving first and second inputs from said users, and for delivering said presentations to said users includes networks including the Internet.
26. The system of claim 17 wherein said users include data mining management consoles integrated with or connected over said network.
27. The system of claim 17 wherein said users include applications running on a computer, system, or other appliance integrated with or connected over said network.
28. The system of claim 17 wherein said users include thin or thick clients in a client-server or browser-server environment integrated with or connected over said network.
29. The system of claim 17 wherein said first inputs include attributes provided by or associated with said users including identity and locator.
30. The system of claim 17 wherein said second inputs include updates, edits, modifications, deletions, replacements, suspensions, changes, improvements, and optimizations of said data mining models and said presentations.
31. The system of claim 17 wherein said data mining system includes reporting systems.
32. The system of claim 31 wherein said reporting systems include outcome monitoring systems and presentation usage monitoring systems.
33. The system of claims 32 wherein said reporting systems are operable to produce ones of said second inputs.

34. The system of claim 17 wherein said data mining system includes servers, memory devices, processing units, input devices, output devices, display devices, and said users for storing and processing sequences of instructions and for enabling communications with said users.
35. The system of claim 34 wherein said sequences of instructions which when executed cause the method of claims 1 through 4 to be performed.



PRIOR ART
FIG. 1

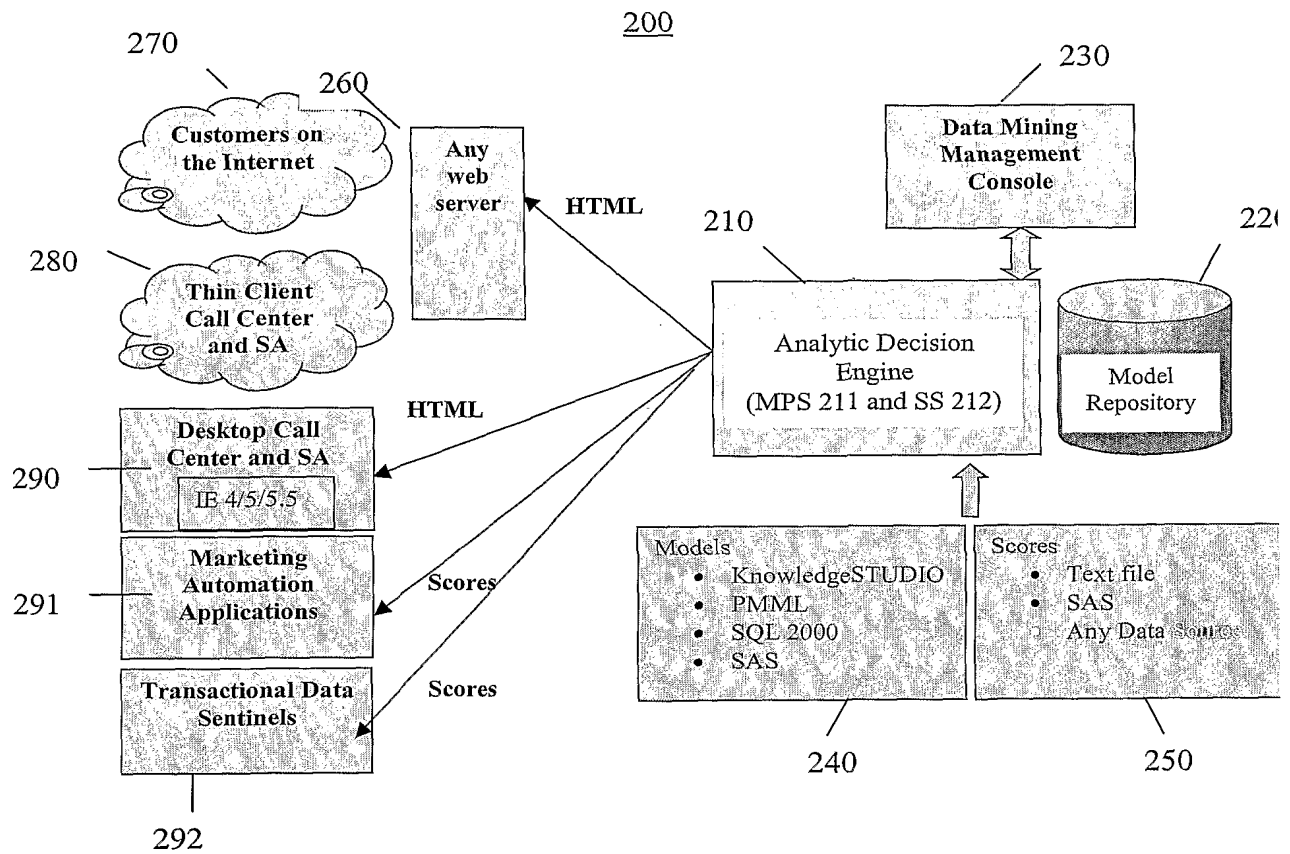


FIG. 2

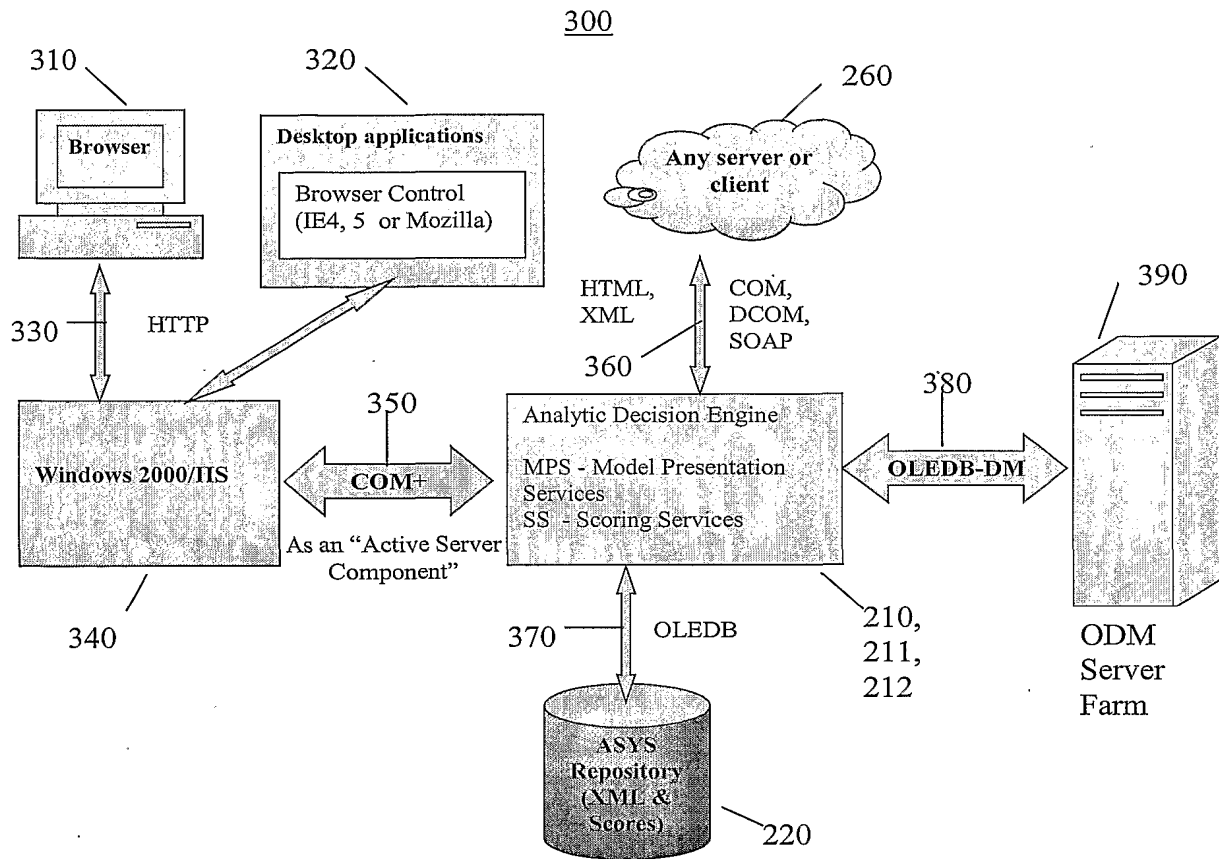


FIG. 3

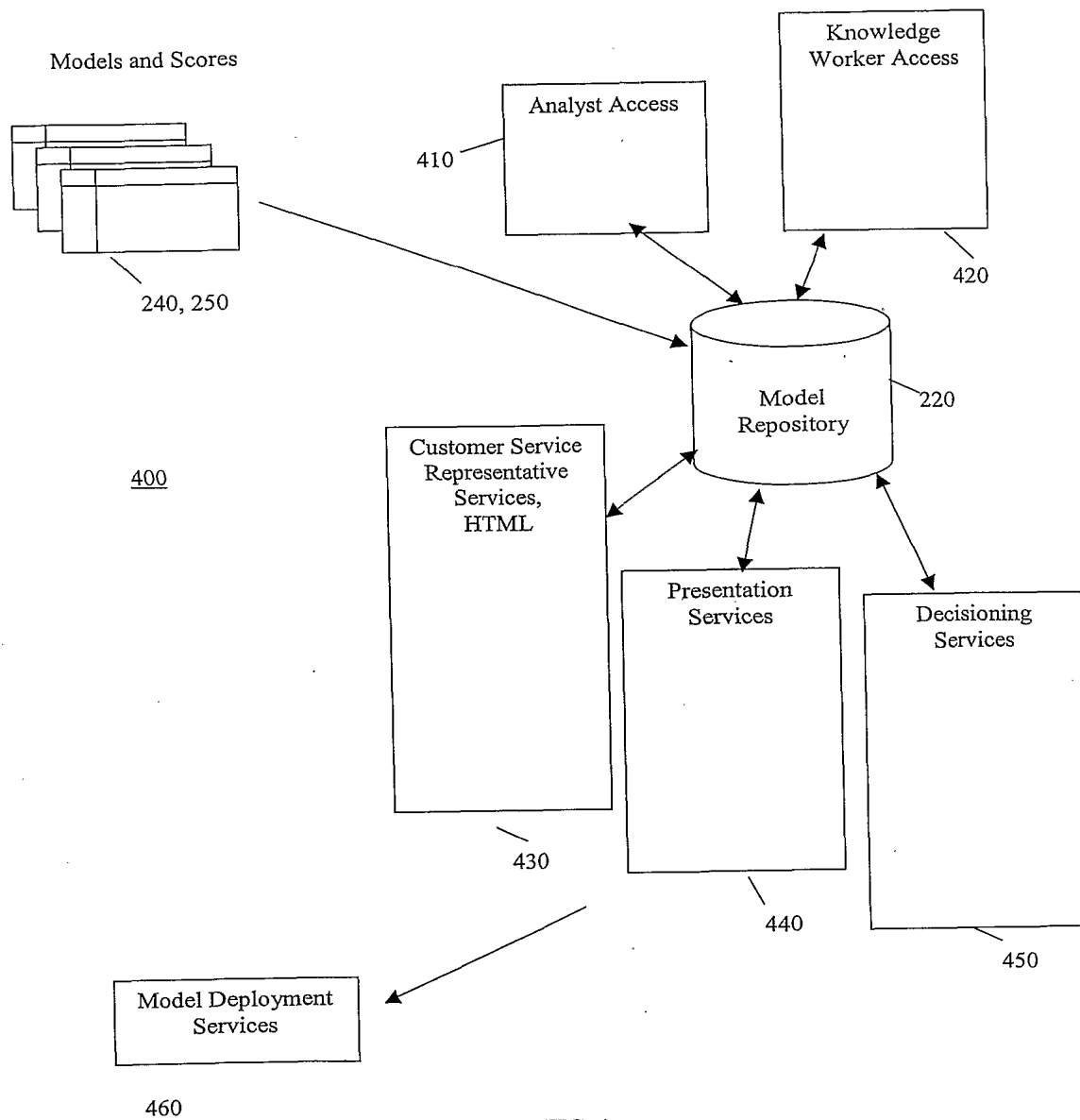
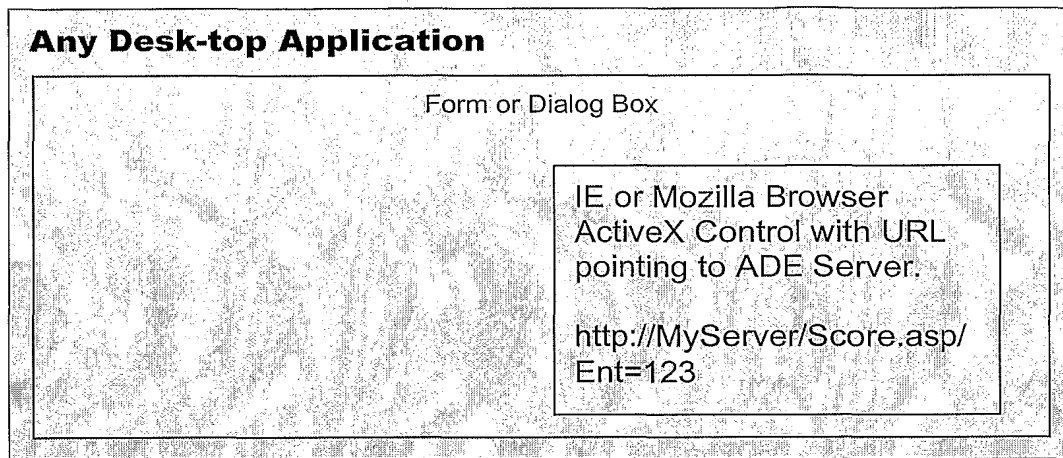
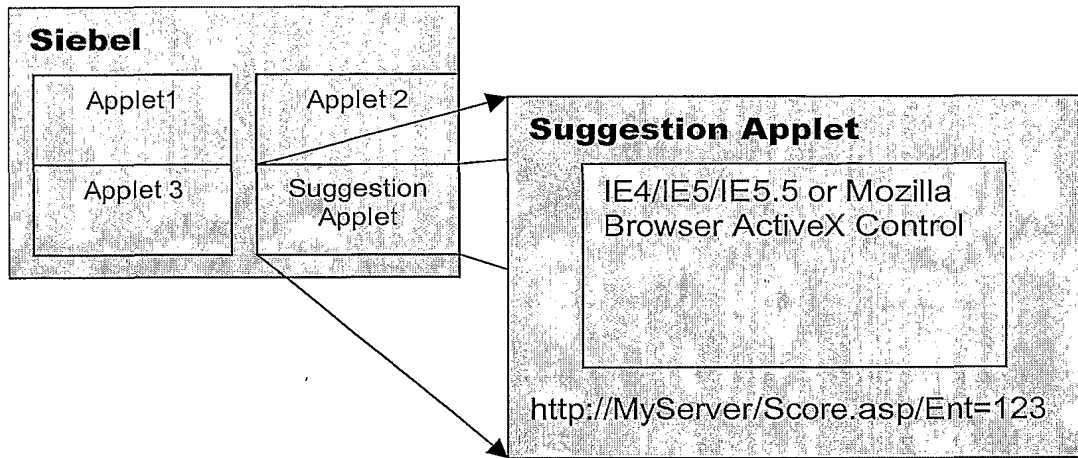


FIG. 4

500a**FIG. 5 (a)**

500b**FIG. 5 (b)**

500c

ANGOSS Knowledge River Siebel Interactron - Call Center Demonstration - Contacts - 7/1/01

File Edit View Screens Go Query Reports Help

History List Threads Contact

Additional

Contacts

Premium Credit Card Offer

We thank you for the call and the opportunity to discuss the premium credit card offer. We hope you are satisfied with the service and the interest rate. Would you like to hear more about this offer?

Would you like to hear more or talk to a specialist?

Would like to hear more Submit

Created Date Note Type Note

Last Name Back First Name Jeff MI Account Site

Job Title Work # Home # City State

Fax # Cellular # Zip Country

Assistant Asst # small

New Activity Type Description Due Assigned To Start Time Alarm

Item: 1 of 20+

FIG. 5 (c)

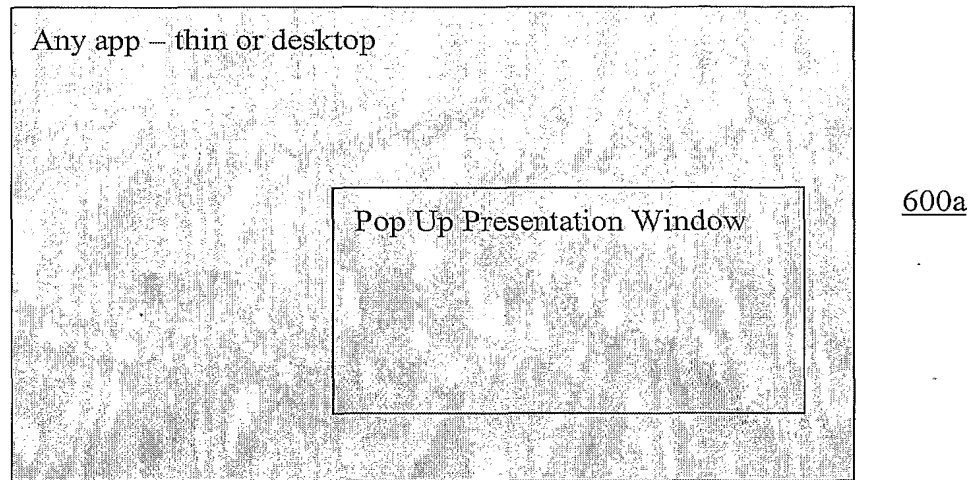
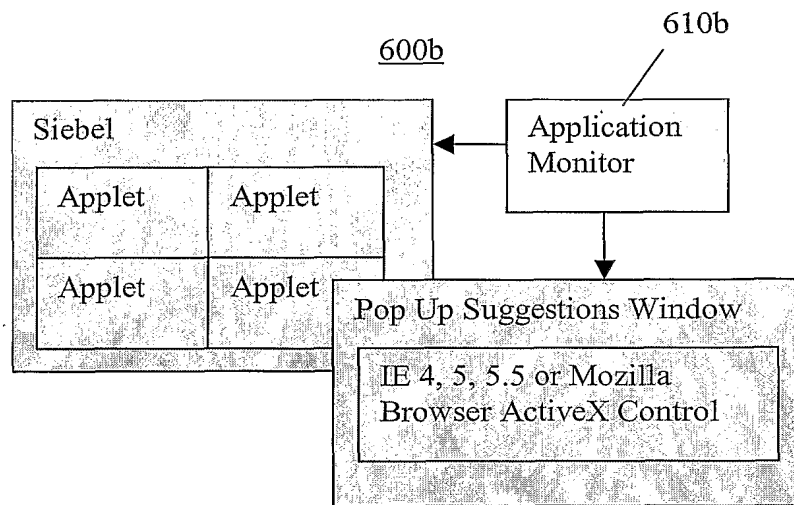


FIG. 6 (a)

**FIG. 6 (b)**

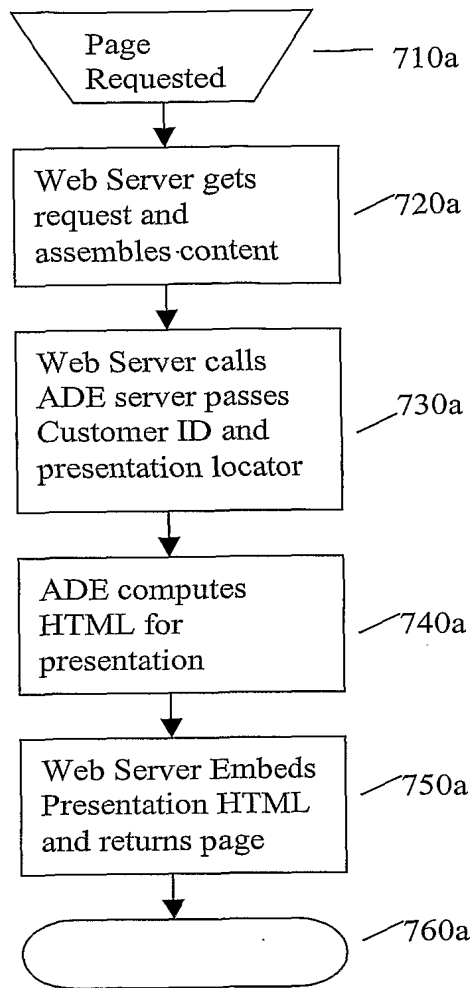


FIG. 7 (a)

Frame 1	Frame 2
Frame 3	Frame 4 Presentation HTML with URL pointing to ADE server. (e.g. http://MyServer/Score.asp/Ent=123)

700b**FIG. 7 (b)**

800

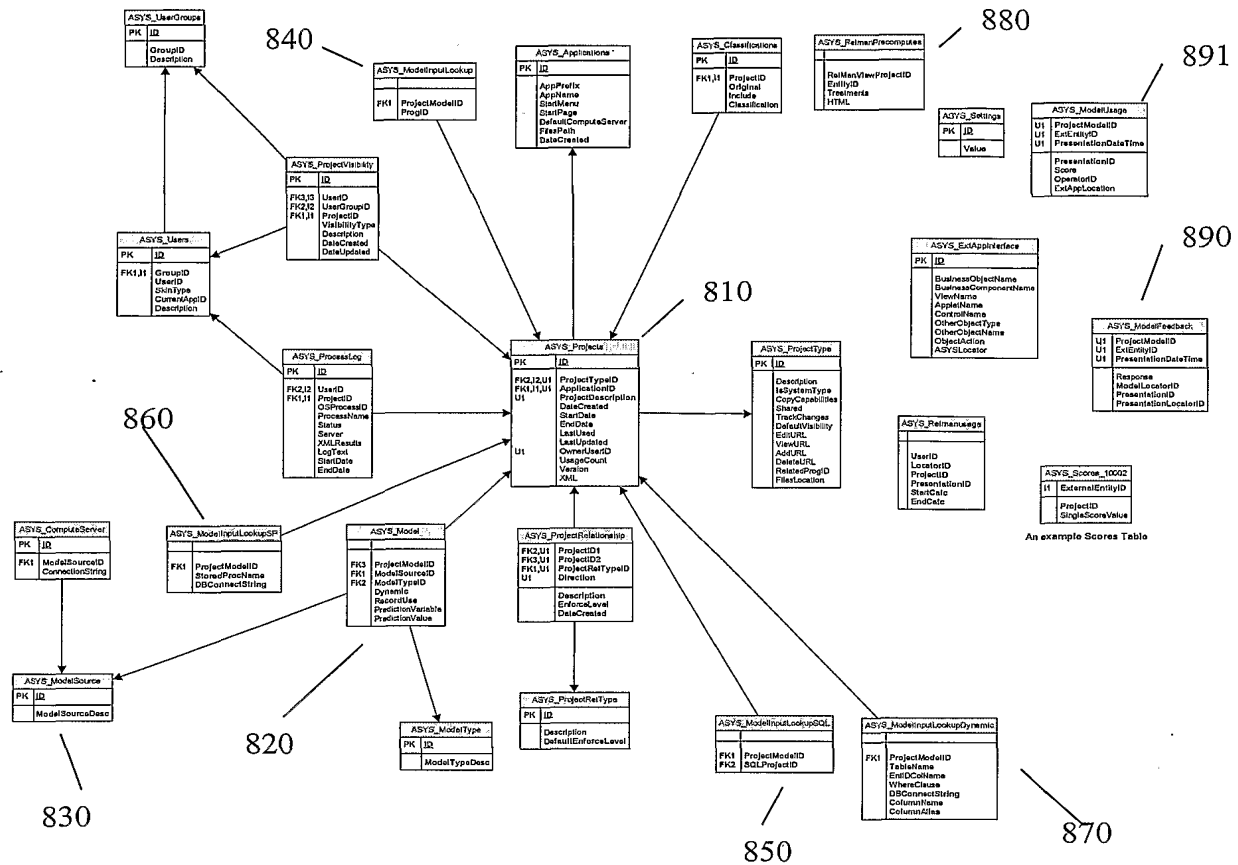


FIG. 8

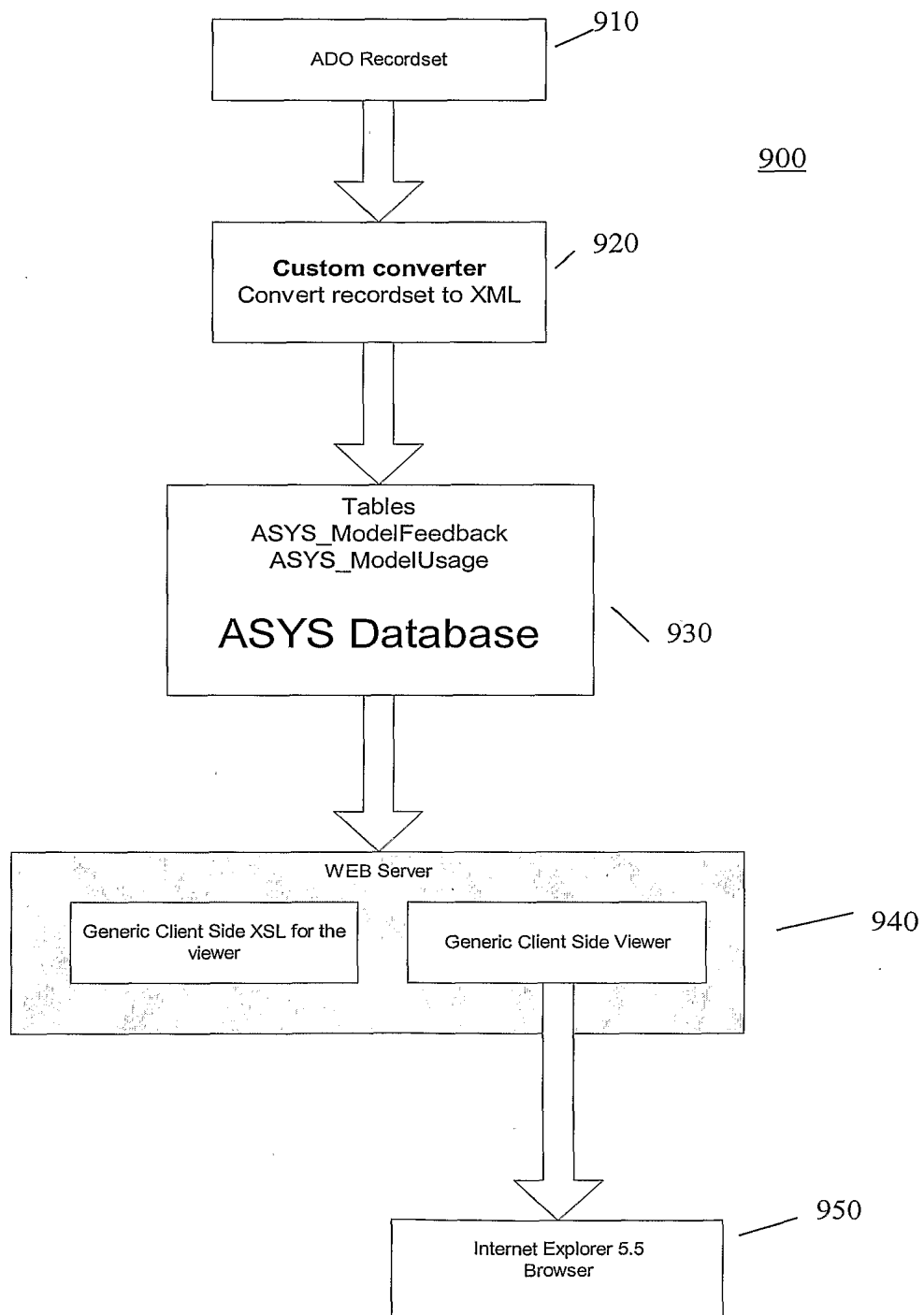


FIG. 9

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
16 January 2003 (16.01.2003)

PCT

(10) International Publication Number
WO 03/005232 A3

(51) International Patent Classification⁷: **G06F 17/60**,
G06N 5/02

(21) International Application Number: PCT/CA02/01044

(22) International Filing Date: 8 July 2002 (08.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/303,036 6 July 2001 (06.07.2001) US

(71) Applicant (for all designated States except US): **ANGOSS SOFTWARE CORPORATION** [CA/CA]; 34 St. Patrick Street, Suite 200, Toronto, Ontario M5T 1V1 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **APPS, Eric** [CA/CA]; 34 St. Patrick Street, Suite 200, Toronto, Ontario M5T 1V1 (CA). **ONO, Ken** [CA/CA]; 34 St. Patrick Street, Suite 200, Toronto, Ontario M5T 1V1 (CA).

(74) Agent: **OGILVY RENAULT**; Kevin Pillay, 1981 McGill College Avenue, Suite 1600, Montreal, Quebec H3A 2Y3 (CA).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

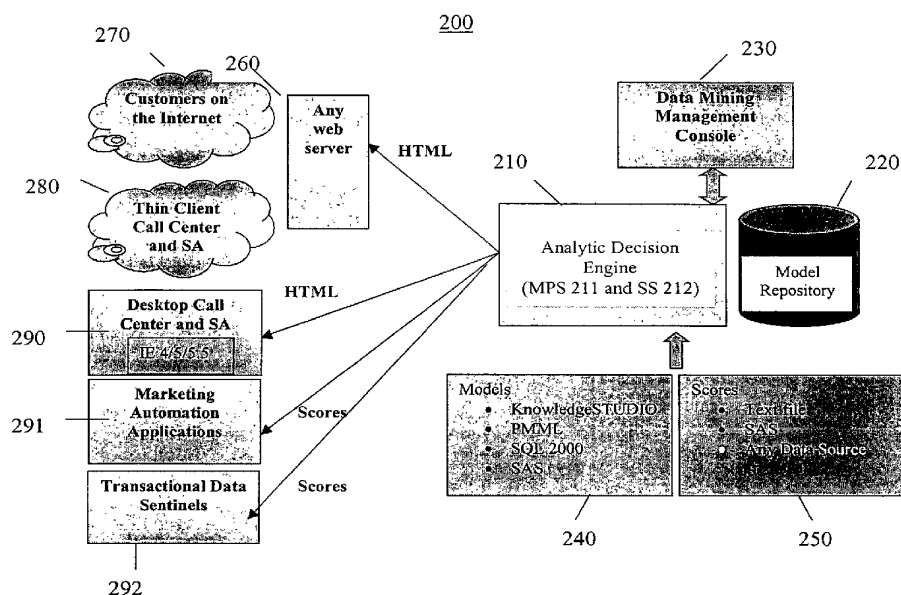
Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:
27 November 2003

[Continued on next page]

(54) Title: A METHOD AND SYSTEM FOR THE VISUAL PRESENTATION OF DATA MINING MODELS



(57) Abstract: A method for delivering presentations associated with data mining models. The method includes the steps of: a) storing the data mining models and the presentations in a repository in a data mining system; b) associating ones of the data mining models with ones of the presentations through references within ones of the presentations; c) receiving a request from a user for a presentation; d) processing the request to determine an appropriate presentation corresponding to the request; the request including a customer identification; the request including an event identification; and, e) delivering the appropriate presentation to the user.



WO 03/005232 A3



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

Interna	Application No
PCT/CA	02/01044

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 7 G06F17/60 G06N5/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 IPC 7 G06F G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 236 978 B1 (TUZHILIN ALEXANDER S) 22 May 2001 (2001-05-22) column 1, line 36 - line 50 column 10, line 50 -column 12, line 24 ---	1-35
X	WO 01 33464 A (SOLOMON NEAL E) 10 May 2001 (2001-05-10) page 21, line 11 -page 26, line 8; figures 9A,10B,19B ----- -/--	1-35

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

1 October 2003

Date of mailing of the international search report

09/10/2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+31-70) 340-3016

Authorized officer

I. SZÁNTÓ

INTERNATIONAL SEARCH REPORT

Intern: " " Application No

PCT/CA 02/01044

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MANGANARIS S ET AL: "A data mining analysis of RTID alarms" COMPUTER NETWORKS, ELSEVIER SCIENCE PUBLISHERS B.V., AMSTERDAM, NL, vol. 34, no. 4, October 2000 (2000-10), pages 571-577, XP004304738 ISSN: 1389-1286 page 572, left-hand column -page 575, left-hand column; figure 1 -----	1-35
X	WO 99 13427 A (HERRINGTON CHERYL ;ARKEL HANS VAN (US); CURTIS TERRIL J (US); GAVA) 18 March 1999 (1999-03-18) page 17, line 26 -page 24, line 11 page 55, line 13 - line 15 -----	1-35
A	US 6 189 005 B1 (SARAWAGI SUNITA ET AL) 13 February 2001 (2001-02-13) column 1, line 1 -column 2, line 28 -----	1-35
A	CHEN M-S ET AL: "DATA MINING: AN OVERVIEW FROM THE DATABASE PERSPECTIVE" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE INC. NEW YORK, US, vol. 8, no. 6, December 1996 (1996-12), pages 866-883, XP001077778 ISSN: 1041-4347 page 866, right-hand column page 869, left-hand column -page 872, right-hand column -----	1-35

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internal Application No

PCT/CA 02/01044

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 6236978	B1	22-05-2001	CA 2309940 A1	27-05-1999
			EP 1029304 A1	23-08-2000
			JP 2002517027 T	11-06-2002
			WO 9926180 A1	27-05-1999
<hr/>				
WO 0133464	A	10-05-2001	AU 1456301 A	14-05-2001
			GB 2369705 A	05-06-2002
			US 2002069134 A1	06-06-2002
			US 2002046157 A1	18-04-2002
			US 2002055903 A1	09-05-2002
			WO 0133464 A1	10-05-2001
			US 2003074301 A1	17-04-2003
<hr/>				
WO 9913427	A	18-03-1999	US 6601048 B1	29-07-2003
			AU 9386598 A	29-03-1999
			CA 2303107 A1	18-03-1999
			EP 1016024 A2	05-07-2000
			JP 2001516107 T	25-09-2001
			WO 9913427 A2	18-03-1999
<hr/>				
US 6189005	B1	13-02-2001	NONE	
<hr/>				